UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/727,169	12/02/2003	Michael L. Kazar	SPIN-5	5977

7590 10/06/2006

Ansel M. Schwartz
Suite 304
201 N. Craig Street
Pittsburgh, PA 15213

EXAMINER	
DOAN, DUC T	
ART UNIT	**PAPER NUMBER**
2188	

DATE MAILED: 10/06/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/727,169	KAZAR ET AL.	
	Examiner	Art Unit	
	Duc T. Doan	2188	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *12 July 2006*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1 and 3-41* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☒ Claim(s) *36 and 37* is/are allowed.

6)☒ Claim(s) *1,3-35 and 38-41* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *12/02/03* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☒ Interview Summary (PTO-413) Paper No(s)/Mail Date. [illegible].
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Continued Examination Under 37 CFR 1.114

A request for continued examination under 37 CFR 1.114, including the fee set for in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.1 14, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.1 14. Applicant's submission filed on 7/12/06 has been entered.

Claims 1-39 have been presented for examination in this application. In response to the last office action, claim 2 has been cancelled, claims 40-41 have been added, claims 1,3,14-15 have been amended. As the result, claims 1,3-41 are now pending in this application.

Applicant's arguments filed 7/12/06 have been fully considered with the results as follows:

Claims 1,3-35,38-41 are rejected.

Claims 36-37 are allowed.

Claims 3,38 are objected.

Claim Objections

Claims 3,38 are objected to because of the following informalities:

As in claim 3, the VFS should not be abbreviated for the initial recital in the claims.

As in claim 38, the NFS should not be abbreviated for the initial recital in the claims.

Appropriate correction is required.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1,3-35,38-41 rejected under 35 U.S.C. 103(a) as being unpatentable over Fridella et al (US Pub 2005/0044090); incorporating reference (Vahalia's et al (5893140) and further in view of Carns et al (PVFS: A Parallel File System for Linux Clusters).

As in claim 1, Frieda describes an apparatus for data storage comprising: a cluster of NFS servers, each server having network ports for incoming file system requests and cluster traffic between servers (Fridella's Fig 1: #110 network file sever, paragraph 27 clustering of data movers #115-#117 provides parallelism as front end to arrays of disks Fig 1: #114 disks);

each server has a network element and a disk element (Frieda's Fig 1, each server #115 must include a network interface element, and a disk/file system interface element; paragraph 28 further discloses each server can be a primary data mover for a disk/file system element; Similarly Friedella teaches by incorporating Vahalia's reference, a network of servers, each server acts a front end (corresponding to the claim's front end network element) for its backend storage disk element such that data in disks can be provided by these servers to clients in parallel manner (Vahalia's Fig 2, Fig 3 show data blocks which are stripped across disks of the disk array

are provided to clients in parallel data stream servers in Fig 2 and Fig 3: channel director/disk director elements; Vahalia's Fig 5, column 7 lines 11-20 further discloses software structures such as common file system, physical file systems in each server, allowing servers to communicating with each other, by network means, to keep track data blocks in file systems which are being distributed across servers);

the servers utilizing a striped file system for storing data (Vahalia's Fig 5, column 7 lines 61-66 further discloses data blocks which are stripped across disks in the disk array are being tracked by physical file systems in each server, and a virtual common file system having maps to map client's file system requests using vnode (virtual node identifier) into the physical node (physical server identifier) in which the data blocks of these file system requests are physically located. Using this virtual node/physical node mapping and logical block address of the client request, the determination of which servers and data blocks serving this client's request can be generated quickly, particularly in the instant of a client's write request, the corresponding data blocks being modified will be immediately flagged (lock flags). This method (i.e mapping vnode/physical node, and maintaining meta data, files attributes of the file systems in primary and secondary movers) will solve the NSF file accessing issue that is normally accesses on a file basis, and thus providing alternative way to access data on block basis (see Vahalia's column 10 lines 1-33 and Friedella's paragraph 40). Therefore, Fridellla clearly teaches a method of data blocks in file systems being stored across disks in the disk array, accessing these blocks in an efficiently parallel manner by multiple secondary movers, and by simply tracking "locking" precisely data blocks being services in these secondary movers. Fridella and Vahalia do not expressly use the word "stripped". However, in a similar manner, Carns discloses a parallel

virtual file system, PVFS, in which data blocks of the file system can be stripped across the I/O nodes/disks (see Carns's Fig 1, section 3.1 second and third paragraphs, trip size is 65 Kbytes). It would have been obvious to one of ordinary skill in the art at the time of invention to include the stripping data blocks of PVFS file system as suggested by Carns in Fridella's system thereby further increasing system data throughput because multiple servers I/O nodes can access data blocks in parallel read/write operations (Carns's section 3.1 paragraph 4).

As in claims 3, the claim recites wherein each disk element has a virtual file system with the virtual file system of each disk element together forming a striped VFS (Vahalia's Fig 5, column 10 lines 34-36 discloses that client can issues multiples requests to multiple servers for data blocks belonged to files, because data blocks are being stripped across these servers/disks. (distributed loading to servers, granularity down to logical data block level).

As in claim 4, the claim recites wherein all disk elements for a virtual file system act as meta-data servers (Fridella's paragraph 28 discloses that each file system is managed by one data mover, primary mover, (corresponding to instant's application meta data server). And the role of being a primary mover can be assigned to any server, for example, Fridella's Fig 1 shows server #111 being a primary mover for file system #121, and server #117 being the primary mover for file system #122).

As in claim 5, Fridella discloses wherein a file has attributes and each server for each file maintains a caching element that stores a last known version of the file attributes and ranges of modification time and change time values for assignment to write operation results (Fridella's paragraph 32 describes file attribute are cached in both primary mover and secondary movers;

Fridella's paragraph 38 further describes the secondary movers maintaining/storing the latest time, using a local value m which obtained from the primary clock and value of a local timer/counter for ranges of modification time values of requests being received in these secondary movers).

As in claim 6, the claim recites wherein each disk element, which is not the meta-data server for a virtual file system, is an input output secondary (Fridella's paragraph 28 further discloses other servers are assigned to move only data, functioning as secondary data movers).

As in claim 7, Fridella discloses wherein ranges of file modification times or file change times are reserved from the meta-data server by the input output secondary (Fridella's paragraphs 36-38 describes the update time is a function of the clock time obtained from the clock of the primary mover which including a value m obtained from the primary clock).

As in claim 8, Fridella discloses wherein the modification and change times in the ranges obtained from the meta-data server are issued to operations already queued at the input output secondary. The claim rejected based on the same rationale as of claim 7. Fridella's paragraph 38 further describes that the secondary movers maintain clock time values using timers/counters and issuing these values for requests in each file's range that it has opened "reserved" for asynchronous write access.

As in claim 9, the claim recites wherein modification and change times in the ranges obtained from the meta-data server are issued to operations received during a window of time after the ranges are reserved from the meta-data server by the input output secondary (Friella's paragraph 46 discloses that after a first secondary mover obtaining a reserved time from the primary mover for the first asynchronous write for a range data blocks, while executing

subsequent write requests for these data blocks (and before committing these data blocks), for example, a second secondary mover may obtain another reserved time for a second set of data blocks; due to clock skewing among different servers (Frieda's paragraph 33, clocks of servers are not synchronized), the first secondary mover must compare these time values and adjust its local cached time value accordingly (also see Frieda's paragraph 12, clock time value from the first mover, time interval measured by the secondary mover)).

As in claim 10, the claim recites wherein operations affecting all stripes of a file begin executions first at the meta-data server for a file and then execute at all input output secondaries, such that operations at the input output secondaries wait only for already executing operations that have already finished their communication with the meta-data server. (Fridella's paragraph 11, all secondary movers's write requests must begin by communicating with the primary mover to obtain meta data values including a modified time value at the first asynchronous write request. Subsequent second asynchronous write requests which are directed to each secondary mover, are handled by the secondary movers without the need to communication with the primary movers. As such, these second asynchronous write requests can be understood as "already finished their communication with the primary mover", as claimed.

As in claim 11 the claim recites an apparatus as described in claim 10 wherein operations follow one of at least two locking models, the first of which is to synchronize first with the meta-data server, then begin core execution by synchronizing with other operations executing at the input output secondary, and the second of which is to first synchronize at the meta-data server, and then to synchronize with operations at one or more input output secondaries that have begun core execution at the input output secondaries. The claim rejected based on the same rationale as

in claim 10. The locking models merely indicate means to reserve a range of data bocks for modifications such that integrity of these data blocks is preserved. Friedella's paragraph 11 teaches the first method to reserve data blocks by communicating with the first/primary data mover, and the second method done by the secondary data mover for subsequent second asynchronous data requests, synchronizing/ordering these requests among themselves (client's requests pending in the secondary data mover, waiting for being executed by the secondary mover in an asynchronously manner).

As in claims 12-13, the claims recite wherein the cluster network is connected in a star topology (claim 12; Fridella's Fig 1, any client can access data using any mover in a star topology); wherein the cluster network is a switched Ethernet (claim 13; Fridella's paragraph 27 lines 10-16).

As in claim 14, the claim recites creating a file across a plurality of NFS servers; writing data into the file as strips of the data in the servers, the strips together forming a stripe; reading strips of the data from the servers; and deleting the strips from the servers. The claim rejected based on the same rationale as in the rejection of claims 1 and 3. Fridella's column 10 lines 25-30 further discloses meta data structures in servers capable of keeping track of data blocks that are destaged/upstaged from the disk array. Thus they are being used to maintain data of the trips in servers, or deleting the trips from the servers when data are destaged to disks.

Claim 15 rejected based on the same rationale as in the rejection of claim 6.

Claim 16 rejected based on the same rationale as in the rejection of claim 5.

Claim 17 rejected based on the same rationale as in the rejection of claim 7.

As in claims 18-19, including the step of making a status request by the caching element to the meta-data server to obtain a file's current attributes (claim 18;) wherein the making a status request step includes the step of obtaining modification time and change time ranges from the meta-data server (claim 19). Friedella's paragraph 43, secondary mover sends a request to the primary mover to obtain the current meta data attribute values, FmpGetAtr, and receiving modification time value from the primary mover.

As in claims 20,21,23 the claims recite including the step of queuing file read and file write requests at the input output secondary until the file read and file write requests are admitted by the cache element and complete execution (claim 20; Fridella's paragraph 39 discloses each secondary mover can process client's subsequent second asynchronous write requests. In order to processing the requests in asynchronous manner, a queue must be employed to keep these clients's requests and subsequently executing them latter, in an asynchronous manner); including the step of tracking by the cache element of the file read and file write requests executing for the file and the ranges that are being read or written (claim 21; Fridella's paragraph 39, each secondary data mover tracks the subsequent second asynchronous write requests for the data blocks/range that it has been reserved at the first asynchronous write request); including the step of checking a byte range affected by a file read request to ensure it does not overlap a byte range of any file write requests previously admitted and currently executing (claim 23; Fridella's paragraph 39 clearly suggests the subsequent second asynchronous write requests must be checked using the modification time value to insure these requests on the data blocks are done in proper order as in client's requests);

As in claims 22,24 the claims recite including the step of requesting the cache element move out of invalid node to read mode when a read operation must be executed (claim 22); including the step of requesting, in response to a file write request that the cache element move into a write mode (claim 24). Fridella's Fig 3 #151 to #159 shows a secondary mover receives a file access request that moves it into a corresponding state (read or write accessing mode) and executing sequence of steps to fulfill the request.

As in claim 25, the claim recites including the step of checking with the cache element the byte range affected by the file write request for overlap with any admitted and still executing file read or file write requests. The claim rejected based on the same rationale as in claim 24, that is to protect the integrity of data being updated, each secondary mover must check subsequent second asynchronous requests using the modify time values, so that these requests from clients are processed in proper order as intended by the client.

As in claim 26, Fridella discloses when executing a write request, of allocating a modification time and change time pair from the range of modification times and change times stored in the cache element (Fridella's paragraph 39 discloses when receiving subsequent second asynchronous write requests, the secondary mover keep tracks of modification times of these request, storing these values in its cache).

As in claim 27, the claim recites including the step of checking the head of a queue of pending file read and file write requests to see if a head request can be admitted by the caching element after either a file read or file write request is completed. The claim rejected based on the same rationale as in the rejection of claim 25.

As in claim 28, Fridella discloses including the steps of detecting by the cache element that a file length must be updated in response to a file write request, moving the cache element into exclusive mode; and making a file write status call to the meta-data server to update length attributes of the file (Fridella's paragraph 39 discloses when the secondary server performing the write commit operation, all previous operations are completed, such that the changed data in this cache can be flushed to disk exclusively, the secondary mover also sends the updated attributes associated with these data blocks/length to the primary server (see Friedella's paragraph 45)).

Claim 29 rejected based on the same rationale as in the rejection of claim 5.

Claim 30 rejected based on the same rationale as in the rejection of claim 7.

Claim 31 rejected based on the same rationale as in the rejection of claim 18.

Claim 32 rejected based on the same rationale as in the rejection of claim 19.

Claim 33 rejected based on the same rationale as in the rejection of claim 22.

Claim 34 rejected based on the same rationale as in the rejection of claim 24..

Claim 35 rejected based on the same rationale as in the rejection of claim 28.

As in claim 38, the rationale in the rejection of claim 1 is incorporated herein. The claim recites a method for reading data in a file comprising the steps of: receiving an NFS read request for data in the file at a network element (Fridella's paragraph 27 lines network devices communicate with each other using NFS and CIFS command); determining by the network element which VFS stores at least one strip containing the data (see rationale of claim 1); sending a file read request from the network element to at least one disk element of a plurality of servers storing a strip of the data (Friedella's paragraph 29, client sending access requests to

multiple secondary movers); obtaining current attributes associated with the file by each disk element; reading the strips of the file from each disk element having the strips; and generating a response in regard to the file read request (Obviously, the secondary movers must generated a response and returning the read data to client).

As in claim 39 the claim recites a method for writing data in a file comprising the steps of: receiving an NFS write request for a file at a network element; determining by the network element which VFS is associated with the file; sending a file write request from the network element to at least one disk element of a plurality of servers having a stripe of the VFS; acquiring current attributes associated with the file; and writing a predetermined number of bytes of the data into each VFS strip in succession until all of the data is written into the file. The claim rejected based on the same rationale as in the rejection of claims 19,38.

As in claim 40, Fridella's paragraph 37 discloses the servers are NFS servers (Fridella's paragraph 40 further discloses that the primary and secondary movers, having proper maps, meta data structures such as vnode/physical node mappings, logical data blocks to physical data block mapping (see Vahalia's column 10 lines 1-33). By using these meta data structures, the data movers can easily operate as NFS servers).

As in claim 41, the claim recites identifying a disk element for a virtual file system of an NFS server as a meta-data server as a meta-data server and a disk elements for the NFS servers which are not identified as the meta-data server as input output secondaries. The claim rejected based on the same rationale as in claim 40. Fridella's Fig 1, paragraph 28 discloses a mover #155 is designated as primary mover for a file system A (corresponding to the claim's meta-data

server), while other movers are designated as secondary movers with respect to file system A (corresponding to the claim's input output secondaries).

Response to Arguments

Applicant's arguments in response to the last office action has been fully considered but they are not persuasive. Examiner respectfully traverses Applicant's arguments for the following reasons:

Firstly, applicant argument is mooted in view of new ground of rejections in view of new references and necessitated by Applicant's amendments.

Secondly, regarding Applicant's arguments in pages 16-23

A) Applicant's argues Fridella does not teach the newly amended limitation "each server has a network element and a disk element", see the rationale of claim 1.

B) Applicant's argues Fridella does not teach "the servers utilizing a striped file system" and teach away from this limitation because Fridella requires secondary movers to obtain a lock from the primary data mover..". and somehow this obtaining the lock teaches away from the stripping data blocks across multiple servers/disks. Examiner respectfully disagrees, Fridella's paragraph 38 clearly discloses that the step obtaining a lock or reserving a range of data blocks is only required for the **first** asynchronous write request of these data block, so that the primary data mover can assign the file modification time values for these data blocks to the secondary mover. For all subsequent second asynchronous write requests, the secondary mover does not have to contact the primary mover, it can executes these second asynchronous write requests independently and in parallel manner, thus the data throughput of the system is greatly improved,

since multiple secondary movers can service multiple second asynchronous write requests for data blocks being stripped across disks in a parallel manner.

Examiner notes that the similar procedure is disclosed in the specification's page 25 second paragraph, that is the secondary I/O doing the first spin write request to obtain a ranges of modification time values from the meta data server. It then uses these time values for accessing these data blocks in subsequent 49 spin write requests. Thus only one (the first spin write request) operation is required to contact the meta data server.

Thus, in contrast to Applicant's allegation that Fridella's "secondary servers obtaining a lock or a reservation for every time the secondary data mover want to access the file system..". Fridella clearly teaches the reservation step only required for the first write request, exactly in the same way disclosed in the instant's application.

Allowable Subject Matter

Claims 36,37 are allowed.

Conclusion

When responding to the office action, Applicant is advised to provide the examiner with the line numbers and page numbers in the application and/or references cited to assist examiner to locate the appropriate paragraphs.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Duc T. Doan whose telephone number is 571-272-4171. The examiner can normally be reached on M-F 8:00 AM 05:00 PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Mano Padmanabhan can be reached on 571-272-4210. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Mano Padmanabhan 9/30/06

DD

Mano Padmanabhan

Supervisory Patent Examiner

Art Unit 2188

MANO PADMANABHAN
SUPERVISORY PATENT EXAMINER